A New Genetic Subtype of HIV-1

Thomas Leitner and Jan Albert

Department of Clinical Virology Swedish Institute for Infectious Disease Control, Karolinska Institute S-105 21 Stockholm, Sweden

During a large screening of Swedish individuals presumed to be infected by subtypes other than the common B subtype (unpublished material), two unrelated individuals were found to carry virus of a previously unknown HIV-1 subtype. Blood samples were collected in Sweden from two HIV-1 infected immigrants from Zaire. Sample SE7022 was obtained in December 1993 from a female who was asymptomatic but had low CD4+ lymphocyte counts (184×10^6 cells/l). The first HIV-antibody positive sample from this female was obtained in May 1990, however the epidemiological investigation indicated that she became infected in Zaire between 1981 and 1986. Sample SE7887 was obtained in October 1994 from a man, who was found HIV-antibody positive in August 1994 (he had tested negative in Sweden in January 1993). His CD4+ lymphocyte count was normal (567×10^6 cells/l). Both patients were heterosexual. A thorough epidemiological investigation, including contact tracing, did not reveal any direct contact between the two individuals.

Sequences of the proviral population, corresponding to the entire *gag* p17 region and the *env* V3 domain, were derived directly from each patient's peripheral blood mononuclear cells by PCR amplification. The fragments were sequenced on both strands using magnetic beads as solid support for purification, dye primers and T7 polymerase for sequencing reactions, and automatic sequencing machines (ALF, Pharmacia Biotech, Sweden) for fragment detection and base calling. The sequencing method allows identification and rough quantifications of heterogeneous nucleotide positions, caused by sequence polymorphisms within the investigated virus population [1].

Sequences SE7887 *env* V3 and p17*gag* are deposited under accession numbers L41176 and L41178, SE7022 under L41177 and L41179, respectively.

The sequences were aligned together with published sequences in the HIV database (*env* V3 region: 291 positions and *gag* p17: 383 positions). Phylogenetic analysis was done with several different methods, including neighbor-joining (using modified Kimura model distances) [2], maximum-likelihood (using empirical nucleotide frequencies and a DNAML transition/transversion ratio of 2.0) [2], uniformly weighted parsimony and *a posteriori* asymmetrically weighted parsimony [3,4]. Bootstrap analysis (1000 resamplings) was done with the neighbor-joining method to test the robustness of the phylogenetic inference.

Figure 1 shows an unrooted neighbor-joining tree of the *env* V3 region sequence fragments (modified from [5]). Trees derived with other methods gave similar results (data not shown). The two new sequences were analyzed together with representatives from subtypes A–H as well as unclassified sequences and the potentially new subtype I found in Cyprus [6]. The two sequences, which clustered closely together, were approximately equidistant to all established subtypes. The closest, yet quite distant, sequence was ZR.Z3, an unclassified sequence from an Zairean individual. The cluster of SE7887 and SE7022 was observed in 100% of the bootstrap resamplings. Sequences from the p17*gag* region were analyzed in the same way, and showed a very similar pattern. The sequences of the two new samples formed a monophyletic clade directly to the center of the tree, giving strong support for the findings in the *env* V3 based trees (Figure 2).

In pattern analysis of the deduced amino acid sequences of the *env* V3 region the two new sequences had four rare amino acids in common as compared to consensus sequences of subtypes A–H and the unclassified sequences (U) (for details see [5]). In p17 *gag* the two sequences differed only in 5 positions, of which three were polymorphic non-synonymous mutations and two were non-polymorphic synonymous mutations.

Furthermore, in the tree analysis, we found that some of the unclassified sequences clustered with established subtypes while others did not. Five out of the eight unclassified sequences were located in or around subtype A and H. The unclassified sequences are described in the Los Alamos database, Part III

Analyses [7]. Subtypes A, G, and H were less well defined than other subtypes. Within these subtypes the members showed a higher degree of genetic divergence, which could be seen in the tree as more deeply rooted branches. This was true using all described tree building methods. As new sequences are added to a study we can expect rearrangement of branches, especially with low number of characters. However, bootstrap analysis confirmed the established subtypes, including the new formed by our two sequences, with high values, except for subtypes H (47%) and A for which no value could be derived because of interference of unclassified sequences. Bootstrap values lower than 70% in most cases correspond to a probability of less than 95% that the clade is true [8].

On the basis of the tree and pattern analyses they fulfill current criteria of being classifed as a new HIV-1 subtype. However, the growth in the number of odd sequences indicates that it is becoming increasingly difficult to categorize the HIV-1 sequences by the currently proposed criteria. We start to see sequences that branch off earlier on established subtypes and sequences that fill the space between already established subtypes, suggesting that we may eventually be faced with a continuum of genetic variants. Variants that represent recombinants between genetic subtypes further complicate the picture.

REFERENCES

[1] Leitner T, Halapi E, Scarlatti G, Rossi P, Albert J, Fenyö EM, and Uhlen M: Analysis of heterogeneous viral populations by direct DNA sequencing. *BioTechniques* 1993;**15**:120–126.

[2] Felsenstein J: PHYLIP: Phylogeny inference package. Version 3.52c. University of Washington, Seattle, Washington, 1993.

[3] Swofford DL: PAUP: Phylogenetic analysis using parsimony. Version 3.1.1. Illinois Natural History Survey, Champaign, Illinois, 1991.

[4] Maddison WP and Maddison DR: MacClade: Analysis of phylogeny and character evolution. Version 3.04. Sinauer Associates, Sunderland, Massachusetts, 1992.

[5] Leitner T, Alaeus A, Marquina S, Lilja E, Lidman K, and Albert J: Yet another subtype of HIV type 1? *AIDS Res Hum Retrovirus 1995*;**11**:995–997.

[6] Kostrikis LG, Bagdades E, Cao Y, Zhang L, Dimitriou D, and Ho DD: Genetic analysis of human immunodeficinecy virus type 1 strains from patients in Cyprus: Identification of a new subtype designated subtype I. *J Virol* 1995;**69**:6122–6130.

[7] Myers G, Korber B, Wain-Hobson S, Jeang K-T, Henderson LE, and Pavlakis GN: *Human retroviruses and AIDS: a compilation and analysis of nucleic acid amino acid sequences*. Los Alamos National Laboratory, Los Alamos, New Mexico, 1994.

[8] Hillis DM, and Bull JJ: An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst Biol* 1993;**42**:182–192.

Editors Note: Tree analysis and pattern analysis using VESPA corroborate the uniqueness of these sequences, which the database will provisionally denote subtype J sequences. Full-length sequences of these viruses are expected in 1996. *GM*

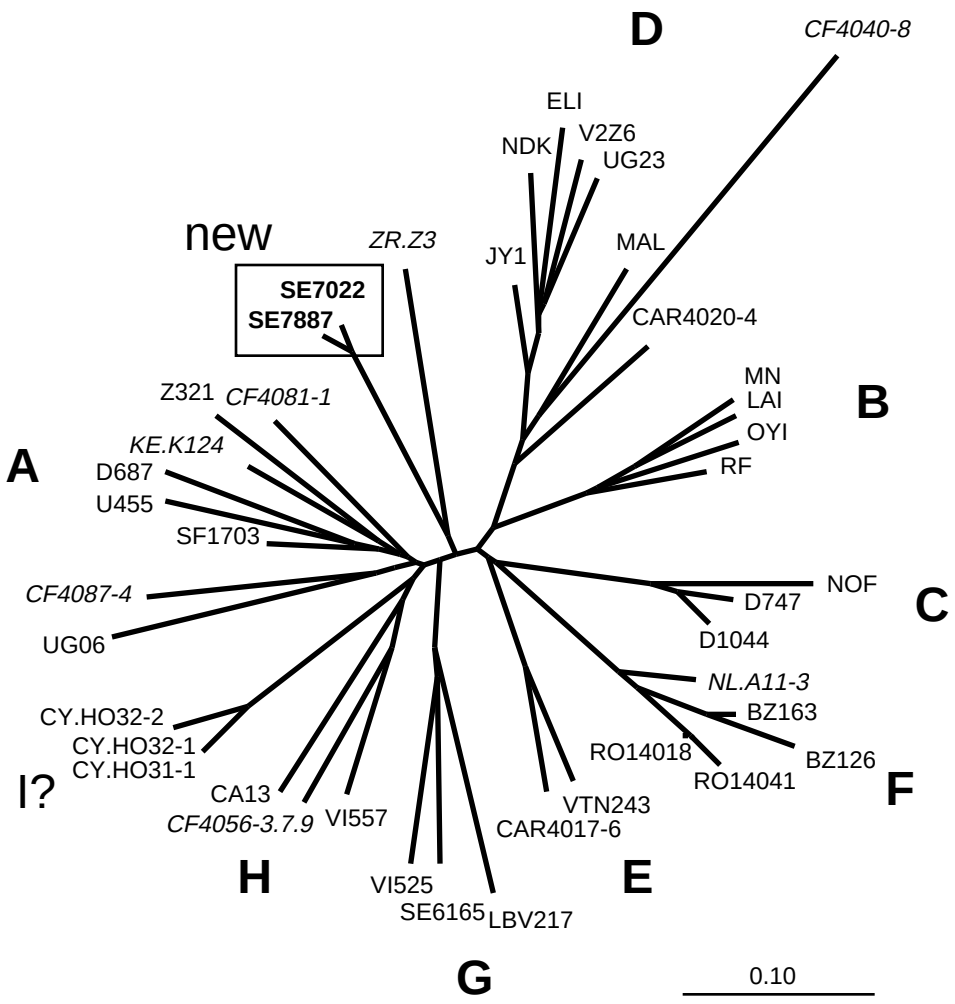

Figure 1. Unrooted neighbor-joining tree of the *env* V3 region sequence fragments (modified from ARHR to include Cyprus sequences). Italic indicates earlier unclassified sequences.

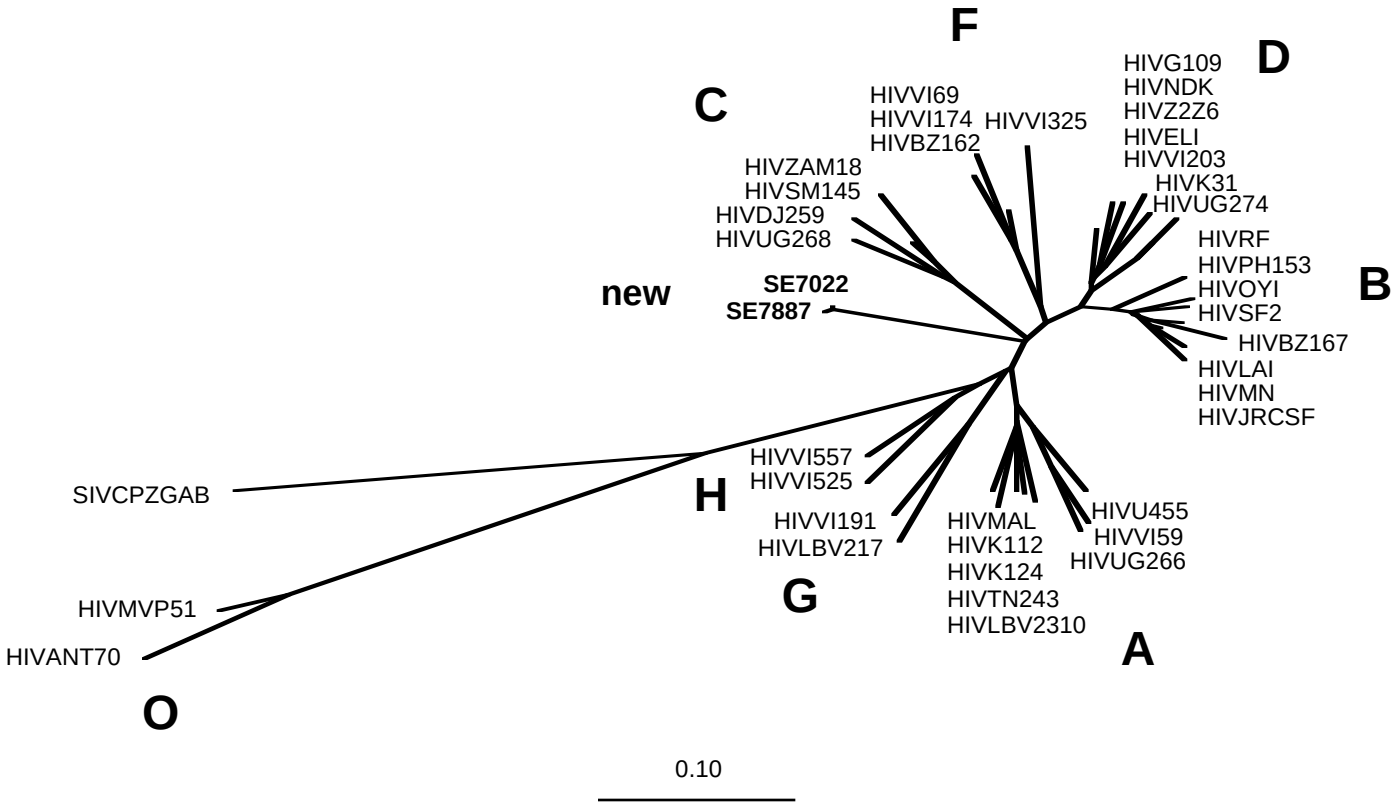

Figure 2 Unrooted neighbor-joining tree of the *gag* p17 region sequence fragments.